Letter to the Editor

International Collaborative Study on Genetic Susceptibility to Environmental Carcinogens[1,2]

Letter
———

Emanuela Taioli[3,4]

Epidemiology Unit, IRCCS, University of Milan, 20122 Milan, Italy

A group of metabolic cancer susceptibility genes exhibit fairly common germline polymorphisms, with frequencies ranging from a few percent to 50% in the general population. These genes encode enzymes that are involved in the metabolism of both exogenous and endogenous toxic substances. In terms of cancer susceptibility, the metabolic gene polymorphisms show a low penetrance and odds ratios for cancer generally on the order of 1.5–2.0. A large number of case-control studies have been conducted on the association between metabolic gene polymorphisms and cancer of various sites (1); however, most of the studies lack adequate statistical power to detect an association, if there is one, due to small sample size (2). A meta-analysis of the published literature in the field confirmed the weak, but significant, associations of certain gene polymorphisms with specific cancer types (1) and also demonstrated the influence of ethnic and or geographic variations in frequencies of these polymorphisms. Potential difficulties with meta-analysis of published results are that not all results might be published and that it is not always possible to compare information

obtained from published studies or to conduct stratified analyses. Because in meta-analysis data on individual subjects are not available, a number of potentially interesting issues cannot be addressed (3).

We have begun a collaborative project to collect and analyze all of the available information on polymorphisms in the genes that metabolize environmental carcinogens, using original data both published and unpublished from laboratories all over the world that are active in this field. The genes to be studied in the first phase of this project include *CYP1A1, GSTT1, GSTM1, NAT2, CYP2E1,* and *CYP2D6.*

Investigators who had previously published case-control studies on genotype for one or more of the genes listed above and a cancer end point up to June 1997 were identified through MEDLINE and contacted by letter. We explained the purposes of the study and asked for published and unpublished original data from case-control studies on gene polymorphism and cancer. Each investigator was asked to provide all original data (without personal identifiers) on all subjects included in his/her study(s). An instruction sheet for submission of data with the variables required and the format was included. A short questionnaire was sent to each investigator to collect information on laboratory methods used to analyze the genotype, study design, response rate, and inclusion criteria. Of sixty-seven investigators contacted, 52 (78%) agreed to participate in the study and 42 (63%) have already sent their data. Data were entered in a data set using Excel after being re-coded in a standard fashion, when necessary. Quality control and logical controls were performed. Gene polymorphisms were all coded according to a standard nomenclature system (4), and submitted data were converted to this system, when necessary. Investigators were asked for explanations of coded data, including genotype data, when necessary. Hard copies of the entered data were sent back to the investigators for further checks. Several other variables not included in the original request were received. These data were included in a separate file, and a list of them was prepared for future use. An Advisory Committee was appointed to help define policies for handling and confidentiality of data and to establish methods for data analysis, data publication, authorship and so on.

The first phase of data collection was concluded on September 1, 1998. As of that date, records from 72 studies for a total of 19,148 subjects, 9,352 controls, and 9,796 cases were received. Half of the studies were conducted using hospital controls, half using healthy subjects. Table 1 shows the distribution of subject cases by cancer site. Basic demographic information, such as age, sex, and race, as well as data on smoking status, were available for most of the subjects. The most frequently analyzed gene in the data set was *GSTM1* (73% of controls, 79% of cases), followed by *CYP1A1* (56% and 60%, respectively; Table 2). In a substantial number of subjects, around 25%, two or three gene polymorphisms were tested simultaneously. Four genes were tested in 17% of the controls and 23% of the cases. A smaller fraction of the data

Table 1 Distribution of cancer cases in the data set by site

Site	n (%)	No. investigators who sent data/no. contacted	No. studies received/no. published (unpublished)
Lung	4819 (49.1)	24/38	29/46 (3)
Skin	997 (10.2)	3/4	3/4 (1)
Breast	807 (8.2)	4/8	4/8 (1)
Bladder	741 (7.6)	4/9	4/9 (2)
Colon	721 (7.4)	8/11	6/9 (2)
Head/neck	656 (7.0)	2/3	2/3 (2)
Prostate	95 (1.0)	2/3	2/3 (1)
Others	956 (9.5)	14/23	5/14 (6)
Total[a]	9792		

[a] Four cases did not have site.

Table 2 Distribution of genetic polymorphisms available in the data set

Gene	Controls n (%)	Cases n (%)
CYP1A1 (*2,*3,*4,*5)	5186 (55.5)	5839 (59.6)
CYP2E1 (*2,*3,*4,*5,*6)	2430 (26.0)	2727 (27.8)
CYP2D6 (*3,*4A,*4B,*4C,*5,*6,*10,*16,*2xN, *2x2,*4x2)	2828 (30.0)	3997 (40.8)
GSTM1	6832 (73.1)	7735 (79.0)
GSTT1	3316 (35.5)	4871 (49.7)
NAT2 (*4,*5A,*5B,*5C,*6A,*6B,*7A,*7B,*14A)	2375 (25.4)	1734 (17.7)
Multiple genes[a]		
Two genes	2347 (25.0)	2476 (25.3)
Three genes	2555 (27.0)	1670 (17.1)
Four genes	1590 (17.0)	2274 (23.2)
Five or more genes	433 (4.7)	1111 (11.3)

[a] These refer to any combination of the above genes.

included information on five or six genes tested simultaneously in the same subject.

The collected data are available to all investigators who submitted data, who may submit proposals for use of the data set to test hypotheses and analyze the data with respect to particular genes, cancers, and other factors. The Advisory Committee reviews the proposals and creates working groups according to common interests. Each investigator who contributes data used in a particular analysis will be given the opportunity to participate in the analysis and will be included among the authors, regardless of their participation in the analysis. Several such analyses are currently in progress. Regular reviews of the published literature are performed every 6 months, and authors of new publications are contacted and asked to furnish their data to maintain this data set as an ongoing system of data collection. In the future, additional genes may be added to the data set as new research directions warrant.

This is the first effort to pool individual data from a large number of epidemiological studies involving metabolic gene polymorphisms as markers of cancer susceptibility. This area of research has produced numerous publications, but many of the results are inconclusive for several reasons, among which are the weak association between each metabolic gene polymorphism and cancer, the small sample size of the studies, and the ethnic heterogeneity of the populations included. Studies on large numbers of cases and testing of multiple gene polymorphisms at once are necessary to better understand the role of metabolic genes in cancer development. This study is an attempt to create a large data set, including several thousand cancer cases and controls from all over the world, where several genes have been tested. The data set can then be used to answer simple questions, such as the level of association of a specific gene polymorphism with a specific cancer, as well as more complex issues such as interactions between different genetic factors and between genetic and other factors (5). In addition, we will compare the results of our pooled analysis with those obtained by classical meta-analysis, to assess the determinants of participation in our project, and possible bias in publication of data (6). One of the major aims of this project is to continue to increase the number of subjects, genes, and other relevant information to expand the number of testable hypotheses. This pooled data analysis project should prove useful not only to confirm associations between gene polymorphisms and cancer that have been suggested in the past, but also and mainly to suggest new working hypotheses related to individual susceptibility to the carcinogenic effects of environmental agents.

References

1. d'Errico, A., Taioli, E., Chen, X., and Vineis, P. Genetic polymorphisms and the risk of cancer: a review of the literature. Biomarkers, *1:* 149–173, 1996.

2. Pearce, N., and Boffetta, P. General issues for study design and analysis in the use of biomarkers in cancer epidemiology. IARC Sci. Publ., *142:* 47–57, 1997.

3. Davey Smith, G., and Egger, M. Meta-analysis—Unresolved issues and future developments. Br. Med. J., *316:* 221–225, 1998.

4. Garte, S., and Crosti, F. A nomenclature system for metabolic gene polymorphisms. *In:* P. Vineis, M. Lang, N. Caporaso, J. Cuzick, and P. Boffetta (eds.). Metabolic Polymorphisms and Cancer. Lyon, France: IARC Publications, in press.

5. Taioli, E., Zocchetti, C., and Garte, S. J. Models of interaction between metabolic genes and environmental exposure in cancer susceptibility. Environ. Health Perspect., *106:* 67–70, 1998.

6. Gordon, I., Boffetta, P., and Demers, P. A. A case study comparing a meta-analysis and a pooled analysis of studies of sinonasal cancer among wood workers. Epidemiology, *9:* 518–524, 1998.

International Collaborative Study on Genetic Susceptibility to Environmental Carcinogens ,

Emanuela Taioli

Updated version	Access the most recent version of this article at: http://cebp.aacrjournals.org/content/8/8/727

Cited articles	This article cites 3 articles, 1 of which you can access for free at: http://cebp.aacrjournals.org/content/8/8/727.full#ref-list-1
Citing articles	This article has been cited by 8 HighWire-hosted articles. Access the articles at: http://cebp.aacrjournals.org/content/8/8/727.full#related-urls

E-mail alerts	Sign up to receive free email-alerts related to this article or journal.
Reprints and Subscriptions	To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.
Permissions	To request permission to re-use all or part of this article, use this link http://cebp.aacrjournals.org/content/8/8/727. Click on "Request Permissions" which will take you to the Copyright Clearance Center's (CCC) Rightslink site.